

EXHIBIT 1

PSYCHO- PHYSICS

Introduction to Its Perceptual, Neural, and Social Prospects

S.S. STEVENS

New Introduction by Lawrence E. Marks

CHAPTER 5

PARTITION SCALES AND PARADOXES

As we have seen, a straightforward procedure for dividing a continuum into equal-appearing intervals is to allow an observer to adjust a series of stimuli to match his conception of equal distances. That was Plateau's procedure when he asked eight artists to paint a middle gray between black and white. The result was a partition scale. But partition scales also come in other varieties. Some of them even turn up in experiments for which the investigator did not intend to produce partitionings. Yet regardless of their origin, all partition scales have one hallmark in common: compared to the magnitude scale, the partition scale is nonlinear. When the partition scale is plotted against the magnitude scale, the result is a curve that is concave downward.

The direction of the curvature of the partition scale is so constant and so independent of the procedures used that it stands as one of the basic invariances of psychophysics. Whenever, given a prothetic continuum, the matching procedure employed is such that the observer is forced out of ratioing and into differencing, so to speak, there results a distorted scale. The distorted scale can usually be described by a power function, so the power function again becomes a convenient descriptor of a stimulus-response relation. But with partition scales the power function has an exponent that is not the actual exponent of the continuum in question. Rather it is a kind of "as if" exponent, a virtual exponent. And as we shall see, the value

of the virtual exponent lies below that of the actual exponent.

CATEGORY SCALE

Perhaps no scale is so well known, so often used, and so generally misunderstood as the ordinary category rating scale. A category scale is created when observers assign a category—such as large, medium, or small—to a series of stimuli. The categories may also be designated by a set of numbers, such as 1 to 7, or by letters of the alphabet, as is customary in the grading of college courses. Interestingly enough, the number of categories used makes relatively little difference to the form of the category scale. If the continuum is prothetic, the category scale exhibits a downward concavity when plotted against the magnitude scale. On metathetic continua, the pure category scale is linear. But since most perceptual continua are prothetic, most category scales share the defect of nonlinearity.

Why the nonlinearity? When he performs category estimation does the observer not assign numbers to stimuli, and is that not what he also does under the method of magnitude estimation? To be sure, the two methods exhibit a superficial resemblance, but they stand poles apart with respect to one crucial feature. Magnitude estimation permits the free assignment of numbers, so that magnitudes, as well as ratios among magnitudes, can be reflected in the numbers that the observer emits to depict his subjective impression. By contrast, category estimation sets a limit on the numbers, letters, or adjectives that the observer is allowed to use. As soon as that limit is set, the nature of the exercise changes and the game takes on a new character. No longer can the observer

respond to relative magnitudes by placing them in their appropriate place on a ratio scale. He must instead try to spread the limited, finite set of numbers over whatever segment of the perceptual continuum is presented to him. In order to spread the numbers evenly, he tries in effect to subdivide the continuum into equal intervals. In other words, whenever the experimenter limits the numbers, the observer's response lapses perforce into a partitioning operation. A finite set of seven numbers, like a set of seven adjectives, can be used to mark off equal appearing distances, but neither the numbers nor the adjectives can be used to express ratios.

It seems intuitively obvious that a series of adjectives do not lend themselves to the representation of ratios, for clearly there is no sense in which a ratio can be said to exist between the adjectives medium and small, or the adjectives large and small. But it may be objected that between two numbers there does exist a definite ratio. In what sense then is the ratio not useful? It is not useful for the same reason that ratios have no firm meaning on interval scales. The problem can be illustrated in this way. Suppose we add an arbitrary number, 100, let us say, to each of the numbers on a 5-point category scale. Instead of the numbers 1, 2, 3, 4, and 5, we then have 101, 102, 103, 104, and 105. For category scaling the second set of five numbers would serve quite as well as the first set. The smallest stimulus would be assigned 101, the largest 105, and the stimulus that seemed halfway between would be given the number 103. Clearly, however, the ratios among the numbers become very different after a constant such as 100 has been added to each of them. Nevertheless, although ratios have been altered,

differences remain unaffected. Consequently, the middle number can still be assigned to the stimulus that appears to lie at the midpoint.

Many psychophysicists have been surprised to find that the category scale is curved in such a way that the midpoint determined by category estimation does not coincide with the midpoint determined by magnitude estimation. Around that issue there has swirled a series of controversies, usually oriented toward two questions: (1) whether it is in fact true that the observer's category judgment fails to hit the same midpoint as his magnitude judgment, and (2) whether, if the two midpoints do not coincide, there exists any basis for choosing one over the other. The first question can be settled directly by experiment, but the second question involves deeper considerations, such as the kind of measurement scale we may prefer to take as basic to the description of sensory functions, and the predictive power that we demand for the resulting system of measurement.

As we found in the previous chapter, the free cross-modality matching of numbers to sensory continua and the matching of one sensory continuum to another have met the severe test of transitivity. In particular, we have seen that the observer's responses in cross-modality matching can be predicted from the responses he makes in free magnitude estimation. Although the demonstration of cross-modal transitivity and its attendant predictive power has impressed many scientists as decisive evidence for the validity of the unconstrained use of numbers in magnitude estimation, not all investigators appear to have been convinced. An occasional author has preferred to believe that a

better scale is obtained when the observer is denied the free use of numbers and is constrained to use a limited set of numbers. Thus one author has interpreted the nonlinear relation between the category and the magnitude scales as evidence that "magnitude estimation is biased." The author goes so far as to say that category ratings "constitute the true measure of sensation" (Anderson 1972). Such a boldly contrary view, if it were to prevail, would foreclose the measurement of sensation on a ratio scale and relegate sensation measurement to a mere interval scale. The greater power and utility of the ratio scale has been demonstrated in many different ways throughout the scientific enterprise. It seems unlikely that the science of psychophysics would be advanced by the deliberate choice of a weaker form of scale.

Many scientists have found it difficult to believe that the simple category scale can be as curved as actual measurement often shows it to be. My own first encounter with a curved category scale left me in a state of puzzled disbelief. The category scale in question was created in 1953 by a group of acoustical engineers who were concerned with the loudness of truck noise on the highway. A jury of listeners rated each passing truck on a 6-point scale. The noises were recorded and also converted into loudness in sones by means of an early version of the sone scale. The engineers seemed pleased with the results, because the two kinds of measures lined up fairly nicely on a graph. The point scattered rather closely about a straight line, with a correlation equal to 0.94. There was a catch, however, for the coordinates were semilogarithmic: the category ratings had been plotted against log sones. In linear coordinates, of course, the plot of category value versus sone value would follow a curve. Since

the sone scale had been constructed to agree with people's judgments, I wondered what had gone awry.

Shortly after my encounter with that engineering study and the question it created about the sone scale, another study gave a similar result. People working in offices on a military air base were asked to express their opinions regarding the noisiness on a category rating scale. Again the category scale agreed better with the logarithm of the sone values than with the linear values.

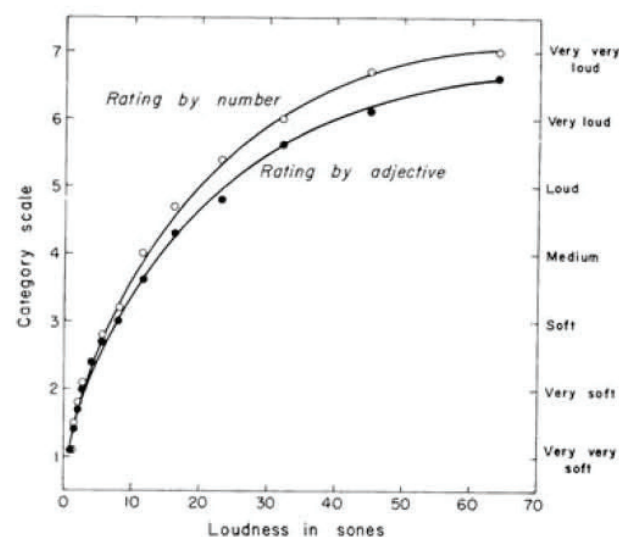


Fig. 47. Results of two experiments showing how the category scale produces a curvilinear plot against the ratio scale of loudness. The stimulus was a white noise that ranged from 40 to 100 decibels in steps of 5 decibels. The logarithmic stimulus spacing tends to increase the curvature. (From Stevens and Galanter 1957, *J. Exp. Psychol.*, 54, 377-411. Copyright 1957 by the American Psychological Association. Reprinted by permission.)

Soon thereafter a colleague and I brought the problem into the laboratory where we analyzed some 70 different category scales on a dozen different continua. On all the prothetic continua, the

category scales were curved. On the metathetic continua, the category scales were straight (Stevens and Galanter 1957).

Figure 47 shows two category scales for loudness, both plotted against the sone scale of loudness. Twenty observers judged the loudness of white noise on a numerical scale from 1 to 7 and on a scale defined by seven adjectives. The faintest and the loudest stimuli were presented at the outset in order to define the two extreme categories. Then the stimuli, spaced at 5 decibel intervals from 40 to 100 decibels, were presented in irregular order, a different order for each observer. The points on the graph represent the mean category rating.

The two functions in Fig. 47 show that it makes little difference to the form of the function whether the observer uses adjectives or numbers. The tendency for the adjectival scale to fall below the numerical scale reflects a local artifact. It happened that some of the observers had been serving in other experiments where the sounds used were much more intense than any used for the category scaling. For those observers the modulus had so changed that the top stimulus did not meet their criterion of "very very loud."

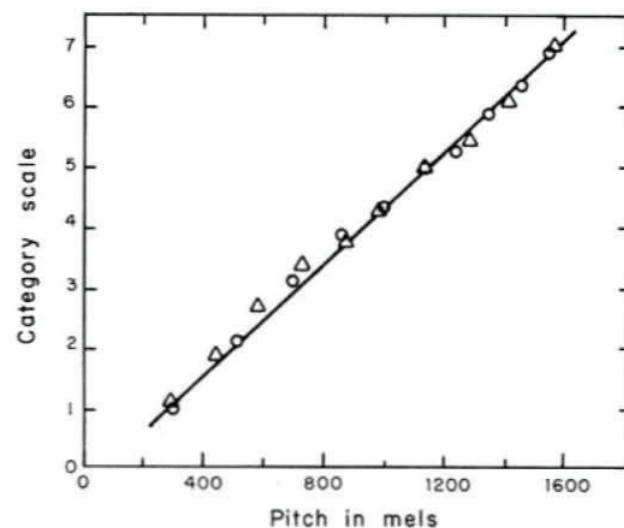


Fig. 48. Category scales for pitch. In two different experiments groups of 10 observers made category judgments of pitch on a 7-point scale. Stimuli were pure tones ranging from 200 to 2000 hertz. (From Stevens and Galanter 1957, *J. Exp. Psychol.*, 54, 377-411. Copyright 1957 by the American Psychological Association. Reprinted by permission.)

The curvature of the prothetic category scale stands in striking contrast to the form of the category scale on a metathetic continuum such as pitch. Figure 48 shows the results of two experiments, in each of which groups of 10 observers made category judgments on a 7-point scale. The listeners judged pure tones ranging from 200 to 2000 hertz, or roughly from well below middle C to three octaves above middle C. The spacing of the tones differed slightly in the two experiments. The average category judgment was plotted against the mel scale of pitch—a ratio scale based on fractionation, equisection, and magnitude estimation. The data in Fig. 48 describe a rather straight line, suggesting that on a metathetic continuum the

category scale is linearly related to the magnitude scale. Apparently, then, category scaling does not necessarily produce a biased, nonlinear outcome, provided its use is restricted to metathetic continua.

CATEGORY SCALES AND THE TIME ERROR

It turns out that certain factors in the experimental procedure can affect the form of a category scale, regardless of the kind of continuum under test. And still other factors that one might suppose would control the form of the category scale have relatively minor effects. Two factors whose effect on the form of the category scale is relatively unimportant are the range of the stimuli presented and the number of categories the observer is permitted to use. Those two factors can vary over rather wide limits without producing a large effect on the form of the category scale.

Since the older literature of psychophysics abounds with descriptions of lifted weight experiments, it is easy to find examples of category scales involving different numbers of categories, as well as stimulus ranges extending from the just discriminable to the widest practicable.

Three illustrative category scales for lifted weights are shown in Fig. 49. The category curves all have about the same form. They would all show slightly greater curvature if, instead of being plotted against the stimulus weight, they were plotted against the subjective weight scale in veks. That follows because the exponent of the veg scale is about 1.5 and therefore concave upward.

Because the curves in Fig. 49 are plotted against grams, they also illustrate the interesting phenomenon known as the "time error," a

constant error discovered long ago by Fechner. On the average, the second of two equal stimuli tends to be judged greater than the first. The time error shows up especially clearly in experiments designed to measure a just noticeable difference on a prothetic continuum. The experiment that produced the triangles in Fig. 49 was such a study. The variable stimuli covered a very short range and the observer was asked to categorize the stimuli as light, intermediate, or heavy.